Cloning and expression in *Bacillus subtilis* of the *npr* gene from *Bacillus thermoproteolyticus* Rokko coding for the thermostable metalloprotease thermolysin

Michael J. O'DONOHUE,*† Bernard P. ROQUES and Ann BEAUMONT

Laboratoire de Pharmacochimie Moléculaire et Structurale, CNRS URA D1500, INSERM U266, Faculté de Pharmacie, Université Paris V, 4 Avenue de l'Observatoire, 75270 Paris, Cédex 06, France

We report the isolation, cloning and expression, in *Bacillus subtilis*, of the gene coding for thermolysin, a thermostable metalloprotease which is produced by *Bacillus thermoproteolyticus* Rokko. The nucleotide sequence has revealed that, like neutral proteases produced by other members of the *Bacillus* species, thermolysin is probably produced as a preproenzyme carrying a typical N-terminal membrane signal sequence. Further, the thermolysin gene shares a strong homology with two other previously cloned genes from two different strains of *Bacillus stearothermophilus*. The sequence of the mature secreted protease, inferred from the DNA sequence, is, with two exceptions, identical with the previously published protein sequence of thermolysin [Titani, Hermodson, Ericsson, Walsh and Neurath (1972) Nature (London) **238**, 35–37]. The exceptions are Asn[37] and Gln[119], originally reported to be Asp and Glu respectively. The biochemical characterization of the secreted recombinant protein shows that it is indistinguishable from the wild-type thermolysin.

INTRODUCTION

Thermolysin (EC 3.4.24.27), a 34.4 kDa thermostable neutral protease from *Bacillus thermoproteolyticus* Rokko (Endo, 1962), is probably the most well-studied member of the zinc endopeptidase family. To date it is one of only four members of this large group of enzymes for which a crystal structure is available, the others being two thermolysin-like enzymes from *Bacillus cereus* (Paupit et al., 1988) and *Pseudomonas aeruginosa* (Thayer et al., 1991) and astacin, a digestive enzyme from the crayfish *Astacus astacus* F (Bode et al., 1992; Gomez-Rüth et al., 1993). Using the data obtained by the co-crystallization of thermolysin with a wide variety of inhibitors, a precise picture of the active site of the enzyme and its probable mechanism of action have been established (reviewed in Matthews, 1988). Thus until recently thermolysin has served as a general model for all zinc endo- and amino-peptidases as, despite having often widely different primary structures, most, if not all of these enzymes, appear to have evolved certain similarities in their active sites and to use the same catalytic mechanism. This approach has already led to the design of highly efficient, selective or 'mixed' inhibitors of mammalian zinc peptidases such as angiotensin-converting enzyme (EC 3.4.15.1), neutral endopeptidase 24.11 (EC 3.4.24.11) and aminopeptidase N (EC 3.4.11.2) (reviewed in Roques and Beaumont, 1990; Roques et al., 1993).

Site-directed mutagenesis of an enzyme, the structure of which has been as intensively studied as that of thermolysin, can be a powerful method for complementing crystallographic data and determining the exact contribution of individual residues in the binding and catalytic processes. With this in mind, the gene from *B. thermoproteolyticus* Rokko, which codes for thermolysin, has now been isolated and cloned. Further, the enzyme has been expressed as an active protein in *Bacillus subtilis*.

EXPERIMENTAL

Materials

All culture media components were Difco products obtained from OSI (Maurepas, France). Oligonucleotides were synthesized by Genosys (Cambridge, U.K.). The Vent exo$^-$ enzyme and the other DNA amplification reagents were from New England Biolabs, being purchased from Ozyme (Montigny-le-Bretonneux, France), while the Sequenase enzyme (version 2) (United States Biochemicals) was purchased from Touzart and Matignon (Vitry-sur Seine, France). [α-^{33}P]dATP and [^{3}H][Leu^5] enkephalin were from NEN-Dupont (Nemours, France). Peptides used for enzymic assays were purchased from Bachem (Bubendorf, Switzerland), while the inhibitor, *N*-[α-rhamnopyranosyl-(oxyhydroxyphosphinyl)]-L-leucyl-L-tryptophan (phosphoramidon), was obtained from Sigma Chemicals (Saint-Quentin Fallavier, France). The other inhibitors, DL-3-mercapto-2-benzylpropanolylglycine (thiorphan) and *N*-[2*R*,2*S*]-3-hydroxyaminocarbonyl-2-benzyl-1-oxopropylglycine (HACBOGly), were synthesized in the laboratory as previously described (Roques et al., 1980; Waksman et al., 1985). Poropak Q was from Waters (Saint-Quentin Fallavier, France) and thermolysin was purchased from Boehringer–Mannheim (Meylan, France). Rainbow molecular markers for SDS/PAGE were from Amersham (Les Ulis, France). All other chemicals were purchased from Sigma Chemicals (Saint-Quentin Fallavier, France) or Prolabo (Paris, France).

Bacterial strains, plasmids, bacteriophage and culture media

The strains used were *Bacillus thermoproteolyticus* Rokko (production of thermolysin, *npr*$^+$) (a generous gift from Dr. R. Sharp, PHLS, Division of Biotechnology, Porton Down, Salisbury,

Abbreviations used: *npr*, neutral protease gene; ORF, open reading frame; thiorphan, DL-3-mercapto-2-benzylpropanolylglycine; HACBOGly, *N*-[2*R*,2*S*]-3-hydroxyaminocarbonyl-2-benzyl-1-oxopropylglycine; phosphoramidon, *N*-[α-rhamnopyransoyl-(oxyhydroxyphosphinyl)]-L-leucyl-L-tryptophan.

* Present address: Laboratoire d'étude des proteines, INRA, Centre de Versailles, Route de St. Cyr, 78026 Versailles Cédex, France.

† To whom correspondence should be sent.

The sequence of the *npr* gene has been submitted to GenBank and has been assigned the accession number X76986.

Wilts., U.K.) and *Bacillus subtilis* DB117 (lacking neutral protease activity, Em^r *his, nprR2, nprE18, aprA3, nfr*$^-$) (Eijsink et al., 1990). For work with the bacteriophage M13mp18 (Yannish-Perron et al., 1982) the *Escherichia coli* strain TG1 was used. The plasmids used for *B. subtilis* were pGE501(Cm^r) (Eijsink et al., 1990) and pGDV1(Cm^r) (Van den Burg et al., 1991). *B. thermoproteolyticus* Rokko was grown in either nutrient or tryptone soya broth at 60 °C. For plates, these media were solidified by the addition of 15 g/l bacteriological agar. *B. subtilis* was propagated in Luria–Bertani broth containing erythromycin or on Tryptose blood agar base (TBAB) plates also containing erythromycin at 37 °C. To screen for protease activity, skimmed milk (1 %, w/v) was included in the solid medium. *B. subtilis* protoplasts were regenerated on DM3 plates (Bron, 1990a). *E. coli* cultures were grown routinely on Luria–Bertani broth at 37 °C. Antibiotics were added to the medium when required, with erythromycin or chloramphenicol being added to a final concentration of 5 μg/ml.

Genomic and plasmid DNA isolation and manipulation

For *B. thermoproteolyticus* Rokko genomic DNA preparation, bacteria were grown in nutrient broth or tryptone soya broth for 16 h at 60 °C. The cells were harvested and washed (50 mM Tris/HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA) and lysed with lysozyme (4 mg/ml), in the presence of 25 % (w/v) sucrose and 50 mM Tris/HCl, pH 8.0. After subsequent treatment with proteinase K (0.4 mg/ml), the mixture was adjusted to 0.66 % (w/v) SDS and 33 mM EDTA, using a concentrated stock solution. This mixture was then incubated at 37 °C until the lysis was completed. The protein was removed by repeated extraction with equal volumes of phenol/chloroform (1:1, v/v) and the nucleic acids were recovered by precipitation with propan-2-ol. Finally, the high-molecular-mass nucleic acids were resuspended in 10 mM Tris/HCl (pH 8.0), 1 mM NaCl, 50 mM EDTA.

Plasmid DNA was isolated from *B. subtilis* by a modified form of the alkaline lysis method described by Bron (1990b). Further purification of large-scale plasmid preparations from *B. subtilis* was achieved using a modified poly(ethylene glycol) precipitation method (Nicoletti and Condorelli, 1993). Extraction and purification of M13mp18 DNA, as well as all other manipulations of DNA, were performed using standard methods described by Maniatis et al. (1989).

Transformation of *B. subtilis* DB117

For plasmid transformation, *B. subtilis* DB117 protoplasts were prepared using a method based on that of Chang and Cohen (1979) which has been described by Bron (1990a).

Oligonucleotides and DNA amplification

The following oligonucleotides were employed to amplify various parts of the thermolysin gene (Figure 1) using a thermal-cycling apparatus (Techne PHC-3) and a heat-stable DNA polymerase (Vent exo$^-$):

APDG1a/b 5′-GTGAGGGAATTCTAYAAYAAYGCITTYT GGAA(C/T)-3′

APDG2 5′-GTGAGGGATCCYTCCCARTCIGGRTTY TTRTT-3′

AP51 5′-GCGCACGTGAAAGATGGC-3′

AP43 5′-CCTCTCCAATTTCCCA-3′

AP29 5′-TACTACTATTTACAAGATAATACGCG-3′

AP38 5′-TCCCACTCCGACAGTTGA-3′

APNT 5′-GGCTCTAGAATCATGACGTTTGGTA-3′

APCT 5′-GTTGAATTCTCCCCTGACATCATTG-3′

where, Y = C or T, R = A or G, and I = inosine, and underlined sequences represent restriction enzyme recognition sites that are introduced into the amplified sequences.

Routinely 350 ng of *B. thermoproteolyticus* Rokko genomic DNA was used for an amplification reaction performed in a 50 μl volume. The final concentration of Mg^{2+} in the reaction, after the addition of the genomic DNA, was adjusted to 2.5 mM using a 100 mM stock solution. Other reaction components were present in the following concentrations: 20 mM Tris/HCl (pH 8.8), deoxyribonucleotides (200 μM each), oligonucleotides (0.25 μM each), 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 0.1 % (v/v) Triton X-100. The thermal-cycling profile used was as follows: 94 °C for 90 s, annealing temperature (an optimized temperature for each pair of oligonucleotides) for 40 s and 72 °C for 180 s (except for the final cycle where the polymerization was allowed to continue for 300 s). Temperature shifts were at 1 °C/s.

DNA sequence determination and analysis

The amplified DNA fragments were cloned either in M13mp18/9 or, in the case of the complete coding sequence, directly in pGDV1. DNA sequencing was performed using the dideoxynucleotide termination method (Sanger et al., 1977) using the Sequenase enzyme (version 2) and [α-^{33}P]dATP. Nucleotide sequences were determined on both strands. The primers used in the amplification reactions served as primers for sequencing, as well as other primers which were specifically designed for this purpose. Nucleic acid and amino acid sequences were analysed using the PC Gene sequence-analysis program (Intelligenetics).

Expression and purification of the extracellular protease

B. subtilis DB117 cells harbouring the plasmid pTLN2 were grown in Luria–Bertani broth containing 5 mM $CaCl_2$ at 37 °C with shaking (180 rev./min.) for 16 h. Thermolysin was purified from the culture supernatant by affinity chromatography using a column (1 cm × 8 cm) of glycyl-D-phenylalanine, coupled to CNBr-activated Sepharose 4B resin following the manufacturer's instructions, and equilibrated in 20 mM sodium acetate, pH 5.5/5 mM $CaCl_2$. The culture supernant, obtained by centrifugation of the culture at 6000 ***g*** for 10 min, was slowly adjusted to pH 5.5 and passed over the column at a flow rate of 6 ml/h. The column was then washed with the equilibration buffer, followed by the equilibration buffer containing 1 M NaCl, and thermolysin was eluted with the 20 mM sodium acetate buffer (pH 5.5) containing 5 mM $CaCl_2$, 2.5 M NaCl and 20 % (v/v) propan-2-ol. Wild-type thermolysin from a commercial source was purified by the same method for comparison. Enzyme purity was assessed by SDS/PAGE using 12 % (w/v) polyacrylamide slab gels. Protein concentration was determined using the method of Bradford (1976).

Assay for enzymic activity

Enzymic activity was normally assayed at 37 °C in a total volume of 100 μl of 50 mM Hepes (pH 7.0) containing 100 mM NaCl

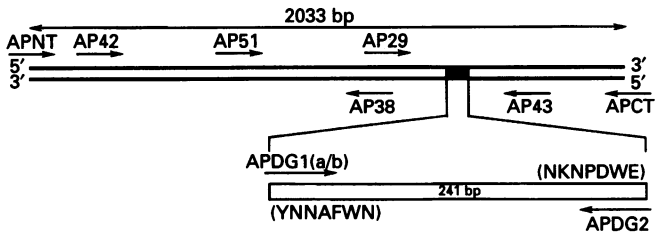

Figure 1 Strategy for the oligonucleotide-directed amplification of the *npr* gene

Various pairs of oligonucleotides were used as primers for DNA amplification in order to specifically amplify regions of the *B. thermoproteolyticus* Rokko *npr* gene. The first pair that was employed consisted of two degenerate oligonucleotides [APDG1(a/b) and APDG2] which were used to amplify a 240 bp region of the *npr* gene (shown as a black box). The positions on the *npr* nucleotide sequence to which these oligonucleotides are complementary is indicated. The respective amino acid sequences in the thermolysin protein which are represented by these degenerate oligonucleotides are indicated above or underneath each of the primers. The approximate positions of the other oligonucleotide primers (which were not degenerate) are represented, with the relevant designation, by arrows which indicate the directional sense of the primer (5′ to 3′).

and 5 mM $CaCl_2$, with 25 nM [^{3}H][Leu^5]enkephalin ([^{3}H]Tyr-Gly-Gly-Phe-Leu) as a substrate. Reactions were stopped by the addition of 10 μl of 0.5 M HCl and the metabolite, [^{3}H]Tyr-Gly-Gly, was isolated (using columns of Poropak Q) as previously described (Vogel and Alstein, 1977). The radioactivity was quantified by liquid-scintillation counting. As the concentration of substrate used for inhibition and pH studies was less than its K_m value for the enzyme, IC_{50} values were taken to be equal to K_i values (Cheng and Prussoff, 1973) and $k_{cat.}/K_m$ was calculated from $k_{cat.}/K_m = v/[E]\cdot[S]$. K_m and $k_{cat.}$ values for [Leu^5]-enkephalin degradation were determined using the substrate over a concentration range of 0.05–4 mM, with 25 nM [^{3}H][Leu^5]-enkephalin included as a tracer. The values were calculated by linear regression analysis, using the program Enzfit (Biosoft).

RESULTS AND DISCUSSION

Cloning of the neutral protease gene (*npr*) in *B. subtilis*

As the DNA sequence for the *npr* gene was unknown two degenerate oligonucleotides were designed; these were APDG1a/b and APDG2, which correspond to residues 110–116 and 180–187 of the thermolysin amino acid sequence respectively. Using these oligonucleotides in a DNA amplification reaction with *B. thermoproteolyticus* Rokko genomic DNA produced a single discrete band, migrating to a position corresponding to a DNA molecule consisting of approx. 260 bp on a 4% (w/v) agarose gel. This DNA was isolated and cloned in M13mp18 and its sequence was determined. Comparison of the sequence with the known sequences of other neutral protease genes revealed a 100% sequence similarity with corresponding sequences in the *npr* M and S genes, which have been previously isolated from *Bacillus stearothermophilus* MK232 and *B. stearothermophilus* TELNE (Kubo and Imanaka, 1988; Nishiya and Imanaka, 1990). Further amplification reactions were then performed using other primers which had been designed on the assumption that the thermolysin gene sequence was highly homologous to those of the *npr* M and S. For the most part, these reactions produced single bands of the expected size, the sequences of which also showed high identity with those of the *npr* M and S genes. Indeed, the less homologous regions are largely confined to the 5′ non-coding region, suggesting that the microbiologically uncharacterized *B. thermoproteolyticus* Rokko may be a strain of *B. stearothermophilus* as previously suggested (Priest, 1989). Finally, the oligonucleotides APNT and APCT were designed and used to amplify a 2033 bp fragment containing the complete *npr* gene and promoter region. The *Xba*I and *Eco*RI sites, introduced in to the amplified fragment by the primers, were then employed to clone the DNA into the *B. subtilis* vector pGDV1. The ligation reaction was used directly to transform *B. subtilis* DB117 protoplasts. Subsequently, several colonies were transferred to TBAB/milk plates to screen for protease activity. One of the lytic halo-producing colonies was selected and plasmid DNA was produced for further sequence analysis and protein purification. This recombinant plasmid was designated pTLN2 (Figure 2).

Nucleotide sequence of the thermolysin gene

The 2033 bp sequence present in pTLN2 is shown in Figure 3. This sequence was corroborated using data obtained from other independently amplified and cloned DNA, eliminating the possibility of experimental artefacts. Analysis of this sequence revealed that the largest open reading frame (ORF) consists of 1644 bp beginning at position 315 and ending at position 1958. Furthermore, in the 5′ flanking sequence putative transcription elements were identified, with potential −35 (TTTTCC) and −10 (TATATT) sequences being found at positions 247 and 268 respectively and a potential Shine–Dalgarno sequence at position 301. These putative signals show a strong resemblance to the promoter of the neutral protease gene, *npr* T (Takagi et al., 1985). In the 3′ non-coding region a possible transcriptional terminator was identified (position 1971 to position 2007).

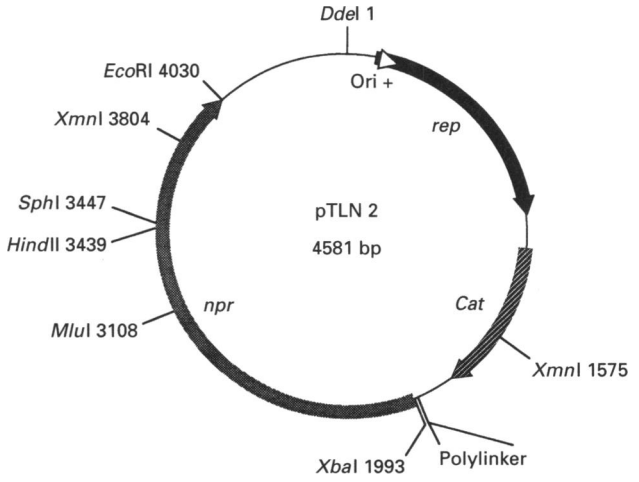

Figure 2 Structure of the plasmid pTLN2

The plasmid is a derivative of pGDV1 containing the complete nucleotide sequence of the *npr* gene from *B. thermoproteolyticus* Rokko which is cloned into the *Eco*RI and *Xba*I sites of the plasmid polylinker. Genes coding for the replicase (rep) and chloramphenicol acetyltransferase (cat) proteins are shaded black and black striped respectively. The positive plasmid origin of replication (ori+) is represented by a white triangle. The neutral protease gene (*npr*) is shaded grey. The position of remaining part of the polylinker region is indicated. Numbering of the plasmid follows the original numbering of Aoki et al. (1987).

```
1     GAATCATGACGTTTGGTATTATGAGTTTACTTGTTCTCATTGGTAGTGACCAAG
55    AACCAAAATATGTGGCAAAAGACGAACATCCGCCTCCAACCATCATCATTGCAG
109   CGAAAGATGAACATCCACCAGCAACGATTATTTGAAGAGGAATAAGCAAAAAGA
163   CAGCTAGTTTTCTAGCTGTCTTTTTTCATGCATAGGAAAATGTGAAAAAAACGT
217   AGGGAATTATCAACTATATCAGACTCTATTTTTCCCAATACAAATACTGTAAAT
                              -35
271   ATTGTGTTAATATTCTAAATACAAAGAATAAAGGAGGATGAAAA ATGAAAATG
      -10                          SD          M  K  M
324   AAAATGAAATTAGCATCGTTTGGTCTTGCAGCAGGACTAGCGGCCCAAGTATTT
       K  M  K  L  A  S  F  G  L  A  A  G  L  A  A  Q  V  F    21
378   TTACCTTACAATGCGCTGGCTTCAACGGAACACGTTACATGGAACCAACAATTT
       L  P  Y  N  A  L  A  S  T  E  H  V  T  W  N  Q  Q  F    39
432   CAAACCCCTCAATTCATCTCCGGTGATCTGCTGAAAGTGAATGGCACATCCCCA
       Q  T  P  Q  F  I  S  G  D  L  L  K  V  N  G  T  S  P    57
486   GAAGAACTCGTCTATCAATATGTTGAAAAAAACGAAAACAAGTTTAAATTTCAT
       E  E  L  V  Y  Q  Y  V  E  K  N  E  N  K  F  K  F  H    75
540   GAAAACGCTAAGGATACTCTACAATTGAAAGAAAAGAAAAATGATAACCTTGGT
       E  N  A  K  D  T  L  Q  L  K  E  K  K  N  D  N  L  G    93
594   TTTACGTTTATGCGCTTCCAACAAACGTATAAAGGGATTCCTGTGTTTGGAGCA
       F  T  F  M  R  F  Q  Q  T  Y  K  G  I  P  V  F  G  A   111
648   GTAGTAACGTCGCACGTGAAAGATGGCACGCTGACGGCGCTATCAGGGACACTG
       V  V  T  S  H  V  K  D  G  T  L  T  A  L  S  G  T  L   129
702   ATTCCGAATTTGGACACGAAAGGATCCTTAAAAAGCGGGAAGAAATTGAGTGAG
       I  P  N  L  D  T  K  G  S  L  K  S  G  K  K  L  S  E   147
756   AAACAAGCGCGTGACATTGCTGAAAAAGATTTAGTGGCAAATGTAACAAAGGAA
       K  Q  A  R  D  I  A  E  K  D  L  V  A  N  V  T  K  E   165
810   GTACCGGAATATGAACAGGGAAAAGACACCGAGTTTGTTGTTTATGTCAATGGG
       V  P  E  Y  E  Q  G  K  D  T  E  F  V  V  Y  V  N  G   183
864   GACGAGGCTTCTTTAGCGTACGTTGTCAATTTAAACTTTTTAACTCCTGAACCA
       D  E  A  S  L  A  Y  V  V  N  L  N  F  L  T  P  E  P   201
918   GGAAACTGGCTGTATATCATTGATGCCGTAGACGGAAAAATTTTAAATAAATTT
       G  N  W  L  Y  I  I  D  A  V  D  G  K  I  L  N  K  F   219
972   AACCAACTTGACGCCGCAAAACCAGGTGATGTGAAGTCGATAACAGGAACATCA
       N  Q  L  D  A  A  K  P  G  D  V  K  S  I  T  G  T  S>  237
1026  ACTGTCGGAGTGGGAAGAGGAGTACTTGGTGATCAAAAAAATATTAATACAACC
       T  V  G  V  G  R  G  V  L  G  D  Q  K  N  I  N  T  T   255
1080  TACTCTACGTACTACTATTTACAAGATAATACGCGTGGAAATGGGATTTTCACG
       Y  S  T  Y  Y  Y  L  Q  D  N  T  R  G  N  G  I  F  T   273
1134  TATGATGCGAAATACCGTACGACATTGCCGGGAAGCTTATGGGCAGATGCAGAT
       Y  D  A  K  Y  R  T  T  L  P  G  S  L  W  A  D  A  D   291
1188  AACCAATTTTTTGCGAGCTATGATGCTCCAGCGGTTGATGCTCATTATTACGCT
       N  Q  F  F  A  S  Y  D  A  P  A  V  D  A  H  Y  Y  A   309
1242  GGTGTGACATATGACTACTATAAAAATGTTCATAACCGTCTCAGTTACGACGGA
       G  V  T  Y  D  Y  Y  K  N  V  H  N  R  L  S  Y  D  G   327
1296  AATAATGCAGCTATTAGATCATCCGTTCATTATAGCCAAGGCTATAATAACGCA
       N  N  A  A  I  R  S  S  V  H  Y  S  Q  G  Y  N  N  A   345
1350  TTTTGGAACGGTTCGCAAATGGTGTATGGCGATGGTGATGGTCAAACATTTATT
       F  W  N  G  S  Q  M  V  Y  G  D  G  D  G  Q  T  F  I   363
1404  CCACTTTCTGGTGGTATTGATGTGGTCGCACATGAGTTAACGCATGCGGTAACC
       P  L  S  G  G  I  D  V  V  A  H  E  L  T  H  A  V  T   381
1458  GATTATACAGCCGGACTCATTTATCAAAACGAATCTGGTGCAATTAATGAGGCA
       D  Y  T  A  G  L  I  Y  Q  N  E  S  G  A  I  N  E  A   399
1512  ATGTCTGATATTTTTGGAACGTTAGTCAAATTTTACGCTAACAAAAATCCAGAT
       I  S  D  I  F  G  T  L  V  E  F  Y  A  N  K  N  P  D   417
1566  TGGGAAATTGGAGAGGATGTGTATACACCTGGTATTTCAGGGGATTCGCTCCGT
       W  E  I  G  E  D  V  Y  T  P  G  I  S  G  D  S  L  R   435
1620  TCGATGTCCGATCCGGCAAAGTATGGTGATCCAGATCACTATTCAAAGCGCTAT
       S  M  S  D  P  A  K  Y  G  D  P  D  H  Y  S  K  R  Y   453
1674  ACAGGCACGCAAGATAATGGCGGGGTTCATATCAATAGCGGAATTATCAACAAA
       T  G  T  Q  D  N  G  G  V  H  I  N  S  G  I  I  N  K   471
1718  GCCGCTTATTTGATTAGCCAAGGCGGTACGCATTACGGTGTGAGTGTTGTCGGA
       A  A  Y  L  I  S  Q  G  G  T  H  Y  G  V  S  V  V  G   489
1782  ATCGGACGCGATAAATTGGGGAAAATTTTCTATCGTGCATTAACGCAATATTTA
       I  G  R  D  K  L  G  K  I  F  Y  R  A  L  T  Q  Y  L   507
1836  ACACCAACGTCCAACTTTAGCCAACTTCGTGCTGCCGCTGTTCAATCAGCCACT
       T  P  T  S  N  F  S  Q  L  R  A  A  A  V  Q  S  A  T   525
1890  GACTTGTACGGTTCGACAAGCCAGGAAGTCGCTTCTGTGAAGCAGGCCTTTGAT
       D  L  Y  G  S  T  S  Q  E  V  A  S  V  K  Q  A  F  D   543
1944  GCGGTAGGGGTGAAA TAAAGTGGTATCTCATCAGTGGGGGATTTTTTCCTCCA

       A  V  G  V  K stop                    TrT              548
1997  CTGATGTTTTGTTTGTGATCAATGATGTCAGGGGAGA
```

Figure 3 Nucleotide sequence of the *npr* gene and the predicted amino acid sequence of the *npr*-encoded protein

The nucleotide sequence of the *npr* gene was determined by DNA sequencing of the plasmid pTLN2. The sequence shown is that of the DNA which was inserted into the *Xba*I and *Eco*RI sites of the plasmid, pGDV1. The putative promoter sequences are underlined and labelled (−35 and −10) as is the putative Shine–Dalgarno sequence (SD). A possible transcriptional terminator (TrT) is underlined. Numbering of the sequence is indicated on the left-hand side. The amino acid sequence (determined by translation of the DNA sequence) of the largest ORF starting at position 315 and ending at position 1958 of the *npr* gene nucleotide sequence is shown. The beginning of the mature thermolysin protein is underlined by a dotted arrow. The two amino acids in the sequence of the mature thermolysin protein which were different to those previously published are double underlined. Numbering of the sequence is indicated on the right-hand side.

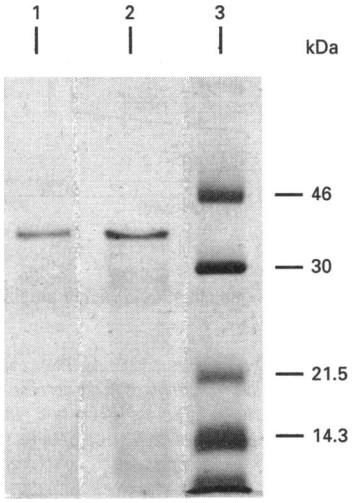

Figure 4 SDS/PAGE of purified wild-type and recombinant thermolysin

Purified wild-type (lane 1) or recombinant (lane 2) thermolysin (5 μg) were run on a SDS/12% polyacrylamide gel and the gels stained with Coomassie Blue. Molecular mass standards are shown in lane 3.

Table 1 K_i values of thermolysin inhibitors for the wild-type and recombinant enzymes

See the Experimental section for details.

	K_i (μM)	
Inhibitor	Wild-type thermolysin	Recombinant enzyme
Phosphoramidon	0.07 ± 0.01	0.072 ± 0.015
Thiorphan	2.2 ± 0.3	2.25 ± 0.41
HACBOGly	5.5 ± 0.7	6.1 ± 0.8
Phenylalanyl-alanine	1000.0 ± 69	1000.0 ± 96

Deduced amino acid sequence of pre-prothermolysin

Translation of the largest ORF of the cloned *npr* gene codes for a large polypeptide consisting of 548 amino acids (Figure 3). The N-terminal sequence displays the distinctive features of a signal peptide with basic residues at the beginning followed by a core of highly hydrophobic amino acids. Lying between this putative signal sequence and the first amino acid of the extracellular protease (Ile[1] of thermolysin) is a 233 amino acid sequence which may represent a prosequence, a feature often found in thermolysin-like proteases from *Bacillus* species (Simonen and Palva, 1993). The ensuing sequence corresponds to the previously reported sequence of thermolysin (Titani et al., 1972), except for two differences: Asp[37] and Glu[119] in the published protein sequence are replaced by Asn (Asn[269] in Figure 3) and Gln

(Gln^{351} in Figure 3) respectively. In addition, residues 301 and 302 of the mature enzyme (residues 533 and 534 in Figure 3), previously reported as Glx-Glx, were found to be Gln-Glu. The apparent differences, which are unlikely to be due to DNA amplification artefacts, probably result from deamination of the amidated residues under the conditions employed for the protein sequence determination.

Characterization of the recombinant protein

Using the glycyl-D-phenylalanine affinity column, approx. 5 mg of recombinant enzyme was recovered/100 ml of culture supernatant. The overall purification factor was not calculated due to component(s) of the culture medium which inhibited enzyme activity. As previously reported for the neutral protease from *B. stearothermophilus* CU21 (Eijsink, 1991), enzymic activity was stable for at least 6 months when stored at −20 °C in the elution buffer. The recombinant enzyme had the same mobility on SDS/PAGE as the wild-type enzyme (Figure 4), the apparent molecular mass of both proteins being 35 kDa, close to the theoretical value of 34.4 kDa for thermolysin. The K_m values for the degradation of [Leu^5]enkephalin were 462 ± 28 μM and 501 ± 41 μM for recombinant and wild-type enzymes respectively, with corresponding $k_{cat.}$ values of 120 ± 12 s^{-1} and 95 ± 15 s^{-1}. The pH dependence of the degradation of 25 nM [^{3}H][Leu]enkephalin was similar for both enzyme preparations, with optimal activity between pH 6.0 and pH 7.0 (results not shown). The inhibition of recombinant and wild-type enzymic activity was compared using representative molecules of different classes of inhibitors for thermolysin (Table 1). The four molecules inhibited recombinant and wild-type thermolysin with similar K_i values.

Taken together, these results confirm that the recombinant protein is identical, in both molecular mass and catalytic properties, to the wild-type enzyme which is produced by *B. thermoproteolyticus* Rokko.

Concluding remarks

Now that the thermolysin gene has been cloned and expressed it should be possible to refine the description of its active site and mechanism of action, using a combination of site-directed mutagenesis and molecular modelling techniques. Furthermore, judicious alteration of active-site residues may help to explain the widely differing substrate preferences of the enzymes in the thermolysin family, thus advancing the goal of having precise models of the active site of each individual enzyme.

Received 1 October 1993/25 November 1993; accepted 16 December 1993

We wish to thank Dr. R. Sharp for his generous gift of *B. thermoproteolyticus* Rokko. Equally we are grateful to Professor G. Venema and Dr. V. Eijsink for their gift of pGDV1 and for their helpful advice and to N. Rousselet and N. Parades for their technical assistance. M.J.O'D. was employed as an associate researcher funded by the CNRS.

REFERENCES

Aoki, T., Nochugi, N., Sasatsu, M. and Kono, M. (1987) Gene **51**, 107–111
Bode, W., Gomez-Rüth, F. X., Huber, R., Zwilling, R. and Stöcker, W. (1992) Nature (London) **358**, 164–167
Bradford, M. M. (1976) Anal. Biochem. **72**, 248–252
Bron, S. (1990a) in Modern Microbiological Methods Series: Molecular Biological Methods for Bacillus (Harwood, C. R. and Cutting, S. M., eds.), pp. 154–155, John Wiley & Sons, Chichester
Bron, S. (1990b) in Modern Microbiological Methods Series: Molecular Biological Methods for Bacillus (Harwood, C. R. and Cutting, S. M., eds.), pp. 142–143, John Wiley & Sons, Chichester
Chang, S. and Cohen, S. N. (1979) Mol. Gen. Genet. **168**, 111–115
Cheng, Y.-C. and Prussoff, W. H. (1973) Biochem. Pharmacol. **22**, 3099–3108
Eijsink, V. G. H. (1991) PhD Thesis, University of Groningen, The Netherlands
Eijsink, V. G. H., Vriend, G., van den Burg, B., Venema, G. and Stulp, B. K. (1990) Protein Eng. **4**, 99–104
Endo, S. J. (1962) J. Ferment. Technol. **40**, 346–353
Gomez-Rüth, F. X., Stöcker, W., Huber, R., Zwilling, R. and Bode, W. (1993) J. Mol. Biol. **229**, 945–964
Kubo, M. and Imanaka, T. (1988) J. Gen. Microbiol. **134**, 1883–1892
Maniatis, T., Fritsch, E. F. and Sambrook, F. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., vol. 1–3 (Maniatis, T., Fritsch, E. F. and Sambrook, F., eds.), Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
Matthews, B. W. (1988) Acc. Chem. Res. **21**, 333–340
Nicoletti, V. G. and Condorelli, D. F. (1993) BioTechniques **14**, 536
Nishaya, Y. and Imanaka, T. (1990) J. Bacteriol. **172**, 4861–4869
Paupit, R. A., Karlsson, R., Picot, D., Jenkins, J. A., Niklaus-Reimer, A.-S. and Jansonius, J. N. (1988) J. Mol. Biol. **199**, 525–537
Priest, F. G. (1989) in Biotechnology Handbooks: Bacillus (Harwood, C. R., ed.), p. 296, Plenum Press, New York
Roques, B. P. and Beaumont, A. (1990) Trends Pharmacol. Sci. **11**, 245–249
Roques, B. P., Fournié-Zaluski, M.-C., Soroca, E., Lecomte, J.-M., Malfroy, B., Llorens, C. and Schwartz, J.-C. (1980) Nature (London) **288**, 286–287
Roques, B. P., Noble, F., Daugé, V., Fournié-Zaluski, M.-C. and Beaumont, A. (1993) Pharmacol. Rev. **45**, 87–146
Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
Simonen, M. and Palva, I. (1993) Microbiol. Rev. **57**, 109–137
Takagi, M., Imanaka, T. and Aiba, S. (1985) J. Bacteriol. **163**, 824–831
Thayer, M. M., Flaherty, K. M. and McKay, D. B. (1991) J. Biol. Chem. **266**, 1864–1871
Titani, K., Hermodson, M. A., Ericsson, L. H., Walsh, K. A. and Neurath, H. (1972) Nature (London) **238**, 35–37
Van den Burg, B., Enequist, H. G., van der Haar, M. E., Eijsink, V. G. H., Stulp, B. K. and Venema, G. (1991) J. Bacteriol. **173**, 4107–4115
Vogel, Z. and Alstein, M. (1977) FEBS Lett. **80**, 332–335
Waksman, G., Bouboutou, R., Devin, J., Besselievre, R., Fournié-Zaluski, M.-C. and Roques, B. P. (1985) Biochem. Biophys. Res. Commun. **131**, 262–268
Yannish-Perron, C., Vierra, J. and Messing, J. (1982) Gene **33**, 103–119